

Predictors of major morbidity and mortality after esophagectomy for esophageal cancer: A Society of Thoracic Surgeons General Thoracic Surgery Database risk adjustment model

Cameron D. Wright, MD,^a John C. Kucharczuk, MD,^b Sean M. O'Brien, PhD,^c Joshua D. Grab, MS,^c and Mark S. Allen, MD^d

Objective: To create a model for perioperative risk of esophagectomy for cancer using the Society of Thoracic Surgeons General Thoracic Database.

Methods: The Society of Thoracic Surgeons General Thoracic Database was queried for all patients treated with esophagectomy for esophageal cancer between January 2002 and December 2007. A multivariable risk model for mortality and major morbidity was constructed.

Results: There were 2315 esophagectomies performed by 73 participating centers. Hospital mortality was 63/2315 (2.7%). Major morbidity (defined as reoperation for bleeding [$n = 12$], anastomotic leak [$n = 261$], pneumonia [$n = 188$], reintubation [$n = 227$], ventilation beyond 48 hours [$n = 71$], or death [$n = 63$]) occurred in 553 patients (24%). Preoperative spirometry was obtained in 923/2315 (40%) of patients. A forced expiratory volume in 1 second $< 60\%$ of predicted was associated with major morbidity ($P = .0044$). Important predictors of major morbidity are: age 75 versus 55 ($P = .005$), black race ($P = .08$), congestive heart failure ($P = .015$), coronary artery disease ($P = .017$), peripheral vascular disease ($P = .009$), hypertension ($P = .029$), insulin-dependent diabetes ($P = .009$), American Society of Anesthesiology rating ($P = .001$), smoking status ($P = .022$), and steroid use ($P = .026$). A strong volume performance relationship was not observed for the composite measure of morbidity and mortality in this patient cohort.

Conclusions: Thoracic surgeons participating in the Society of Thoracic Surgeons General Thoracic Database perform esophagectomy with a low mortality. We identified important predictors of major morbidity and mortality after esophagectomy for esophageal cancer. Volume alone is an inadequate proxy for quality assessment after esophagectomy.

GTS



Earn CME credits at
<http://cme.ctsnetjournals.org>

Despite advances in the management of patients with esophageal cancer, esophagectomy remains an operation with relatively high morbidity and mortality. Pulmonary complications in particular contribute to prolonged hospital stays and poor patient outcomes.^{1,2} Numerous single-institution reports, most with limited data sets, have attempted to identify predictors of morbidity and mortality. Some factors, including weight loss, functional status, smoking status, forced expiratory volume in 1 second (FEV₁), and age are generally thought to influence outcomes,^{3,4} and others, including pre-

operative induction therapy, are more controversial.⁵⁻⁷ To date, however, there has been only 1 large, multi-institutional prospective study of perioperative risk factors for esophagectomy for cancer. Although based on clinical data reporting, this study was limited to patients within the Veterans Administration system, a very select patient population not representative of the general US patient population.⁸ As such, the collected variables were not always applicable to the usual patient population (eg, ascites), and certain important outcome measures (eg, anastomotic leak) were not recorded.

Reports from numerous administrative databases have shown a substantial volume performance relationship for esophagectomy.^{9,10} Many advocate using only volume to judge quality of care. The danger of such a simplistic approach is that some surgeons will be unfairly judged when volume is used as the sole proxy quality measure. This study addresses the need for a morbidity and mortality model for esophagectomy for participants in the Society of Thoracic Surgeons General Thoracic Database.

PATIENTS AND METHODS

Society of Thoracic Surgeons Database

The Society of Thoracic Surgeons (STS) General Thoracic Database (GTDB) was established in 1999 as a voluntary initiative to support the continuous quality improvement efforts of surgeons and hospitals. Participating institutions receive twice-yearly feedback reports that describe each site's results in relation to other database participants. Although the database is not currently audited, all participants sign a contract that requires complete

From the Division of Thoracic Surgery, Massachusetts General Hospital, Boston, Mass.^a

Division of Thoracic Surgery, University of Pennsylvania, Philadelphia, Pa,^b Duke Clinical Research Institute, Duke University, Durham, NC,^c and the Division of General Thoracic Surgery, Mayo Clinic School of Medicine, Rochester, Minn.^d

Presented at the 88th annual meeting of The American Association for Thoracic Surgery, May 10-14, San Diego, Calif.

Received for publication May 7, 2008; revisions received Sept 30, 2008; accepted for publication Nov 16, 2008.

Address for reprints: Dr. Cameron D. Wright, Massachusetts General Hospital, Thoracic Surgery, Blake 1570, 55 Fruit Street, Boston, MA 02114 (E-mail: wright.cameron@mgh.harvard.edu).

J Thorac Cardiovasc Surg 2009;137:587-96
0022-5223/\$36.00

Copyright © 2009 by The American Association for Thoracic Surgery
doi:10.1016/j.jtcvs.2008.11.042

Abbreviations and Acronyms

ASA	= American Society of Anesthesiology
CAD	= coronary artery disease
CHF	= congestive heart failure
COPD	= chronic obstructive pulmonary disease
FEV ₁	= forced expiratory volume in 1 second
GTDB	= General Thoracic Database
NIS	= Nationwide Inpatient Sample
PVD	= peripheral vascular disease
RAR	= risk-adjusted rate
SEER	= Surveillance, Epidemiology, and End Results
STS	= Society of Thoracic Surgeons
VA NSQIP	= Department of Veterans Affairs National Surgical Quality Improvement Program
BMI	= body mass index

reporting of all cases and prohibits selective reporting. Details of the STS GTDB data collection instrument can be found on the STS website.¹¹ Participation in the STS GTDB requires initial institutional review board approval, but subsequent deidentified data analysis for quality improvement purposes does not.

Patient Population

Between January 2002 and December 2007, there were 2391 esophagectomies for primary esophageal cancer entered into the STS GTDB by 75 surgical sites. Data from 2 sites ($n = 49$ operations) were excluded because these sites did not report outcomes consistently during the study period. Also excluded were emergency operations ($n = 14$) and patients younger than 30 years ($n = 6$), as well as cases with missing data for age ($n = 2$) or discharge mortality ($n = 5$). The final study population consisted of 2315 operations from 73 participating sites.

Outcome Definitions

Postoperative events were those defined by the STS GTDB guidelines.¹¹ Hospital mortality is defined as death during the same hospitalization as surgery regardless of timing. We chose hospital mortality rather than 30-day mortality because there were more deaths in the hospital mortality cohort (63 vs only 50 at 3 days) and 7% of sites did not submit 30-day mortality data. The hospital mortality of 2.7% (63/2252) after esophagectomy proved too low to serve as an end point for comparing hospital performance. We therefore decided to analyze a composite morbidity/mortality outcome. Adverse outcome measure selection was on the basis of clinical judgment, literature review, and preliminary data analysis. This outcome was defined as the presence of 1 or more of the following postoperative conditions: bleeding requiring reoperation, anastomotic leak requiring medical or surgical treatment, reintubation, initial ventilation > 48 hours, pneumonia, or death. Twenty-four percent (553/2315) of patients exhibited 1 or more of the defined morbidity events, and this clinically relevant composite outcome measure proved satisfactory to model.

Selection of Covariates

Model variables were identified by reviewing 3 versions of the STS data collection instrument (v1.3, v2.06, v2.07). Because the primary purpose of the model was to adjust for case mix in making hospital comparisons, can-

didate predictor variables were limited to preoperative patient factors that were not directly modifiable by the surgeon or hospital. Variables were excluded from consideration if they were not collected consistently across all versions of the data collection instrument. Remaining variables were screened based on a combination of literature review and informal empirical analysis. Pulmonary functional variables (forced expiratory volume, diffusing capacity for carbon monoxide) were excluded on the basis of excessive (>50%) missing data. Prior cardiothoracic surgery was excluded based on a priori clinical judgment and the lack of significant association with the model end point in univariate analyses. All remaining variables were included in the multivariable analysis and are listed in Table 3. Missing predictor values were managed using a combination of complete case analysis and single and multiple imputation, as described in the Appendix. A sensitivity analysis was conducted to determine whether inferences changed depending on the method of handling missing data.

Multivariable Analysis

Multivariable logistic regression was used to estimate the relationship between patient preoperative characteristics and the outcome of mortality or major morbidity. All covariates were retained in the model and were not added or removed based on a variable selection algorithm. Parameters of the logistic model were estimated using generalized estimating equations methodology to account for statistical dependence between outcomes of patients at the same hospital. Discrimination of the model was assessed by calculating the area under the receiver operating characteristic curve, also known as the C-statistic. The method of bootstrap resampling was used to correct the C-statistic for optimism.

Analysis of Volume Performance Association

To assess the association between STS participant volume and performance, the model described above was subsequently altered to include each participant's average annual volume of esophagectomy procedures. The average annual volume was calculated as 12 times the total number of records from a site divided by the total number of months during which the site submitted data.

Analysis of Hospital Performance Variation

To explore variation in hospital performance, the model described above (without volume) was subsequently refit as a 2-level hierarchical model with nesting of patients within participants. The hierarchical model included the same set of patient factors described above, plus a set of random hospital-specific effects. The hospital-specific effects are interpreted as reflecting underlying differences in performance that systematically increase or decrease risk of all patients at the same hospital. Performance variation was summarized by calculating hospital-specific risk-adjusted rates (RARs) of mortality or major morbidity. The RAR is interpreted as the rate of mortality or major morbidity that would be observed hypothetically for a participant if the participant performed surgery on each patient in the STS database. If the risk-adjusted rate is significantly different from the overall STS database average, this may be interpreted as evidence of good or poor performance. Uncertainty surrounding the estimated RAR was quantified by calculating Bayesian 95% probability intervals. Details of the hierarchical model, including the calculation and interpretation of RARs and probability intervals, are described in the Appendix.

Analysis was performed using S-Plus 6 (Insightful Corp, Seattle, Wash), SAS 9.1 (SAS Institute, Cary, NC), and WinBUGS 1.4.1 (Freeware, <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>) and Imperial College of Science, Technology and Medicine at St Mary's, London).

RESULTS

The average site-specific volume of esophagectomies ranged from 1 to 83 cases per year. Patient characteristics are seen in Table 1. The patients were almost all white

TABLE 1. Distribution of risk factors and frequency of adverse outcomes in study population

Variable	Number of patients		Frequency of mortality/morbidity		P
	n	% of population	n	% of subgroup	
Total	2315	100.0	553	23.9	
Age (y)					
<60	811	35.0	168	20.7	.045
60–64	411	17.8	96	23.4	
65–69	399	17.2	93	23.3	
70–74	324	14.0	79	24.4	
75–80	260	11.2	82	31.5	
>80	110	4.8	35	31.8	
Gender					
Male	1898	82.0	446	23.5	.335
Female	417	18.0	107	25.7	
Race					
White	2104	90.9	500	23.8	.007
Black	63	2.7	24	38.1	
Other	106	4.6	24	22.6	
Missing	42	1.8	5	11.9	
Zubrod score					
0, no symptoms	497	21.5	111	22.3	.086
1, fully ambulatory	1480	63.9	354	23.9	
2, in bed <50%	134	5.8	45	33.6	
3, in bed >50% and <100%	59	2.5	18	30.5	
4, bedridden	8	0.3	2	25.0	
Missing	137	5.9	23	16.8	
ASA					
I	92	4.0	13	14.1	<.001
II	720	31.1	131	18.2	
III	1226	53.0	330	26.9	
IV	129	5.6	47	36.4	
V	2	0.1	0	0.0	
Missing	146	6.3	32	21.9	
Body mass index (kg/m ²)					
<25	648	28.0	160	24.7	.023
25–29	784	33.9	169	21.6	
30–34	445	19.2	114	25.6	
≥35	237	10.2	72	30.4	
Missing	201	8.7	38	18.9	
CHF					
No*	2264	97.8	526	23.2	.006
Yes	51	2.2	27	52.9	
CAD					
No*	1881	81.3	407	21.6	<.001
Yes	434	18.7	146	33.6	
PVD					
No*	2183	94.3	498	22.8	<.001
Yes	132	5.7	55	41.7	
Hypertension					
No*	1177	50.8	236	20.1	.003
Yes	1138	49.2	317	27.9	

TABLE 1. Continued

Variable	Number of patients		Frequency of mortality/morbidity		<i>P</i>
	n	% of population	n	% of subgroup	
Induction therapy					
No*	1299	56.1	319	24.6	.427
Yes	1016	43.9	234	23.0	
Steroids					
No*	2275	98.3	538	23.6	.368
Yes	40	1.7	15	37.5	
Prior cardiothoracic operation					
No*	2084	90.0	484	23.2	.098
Yes	231	10.0	69	29.9	
Diabetes					
No diabetes	1883	81.3	425	22.6	.010
Diabetes, noninsulin	287	12.4	88	30.7	
Diabetes, insulin	86	3.7	32	37.2	
Missing	59	2.5	8	13.6	
Renal function					
No renal insufficiency	2224	96.1	530	23.8	.280
Creatinine ≥ 2	46	2.0	15	32.6	
Dialysis of any type	3	0.1	2	66.7	
Missing	42	1.8	6	14.3	
Cigarette use					
No	579	25.0	115	19.9	.024
Yes	1731	74.8	437	25.2	
Missing	5	0.2	1	20.0	
Stage					
1	494	21.3	116	23.5	.104
2a	442	19.1	122	27.6	
2b	96	4.1	25	26.0	
3	512	22.1	123	24.0	
Missing	771	33.3	167	21.7	
Percent FEV ₁					
<60	89	3.8	32	36.0	<.001
60–79	189	8.2	63	33.3	
≥80	645	27.9	144	22.3	
Missing	1392	60.1	314	22.6	
Percent DLCO					
<60	119	5.1	45	37.8	<.001
60–79	249	10.8	68	27.3	
≥80	351	15.2	76	21.7	
Missing	1596	68.9	364	22.8	

ASA, American Society of Anesthesiology; CAD, coronary artery disease; CHF, congestive heart failure; DLCO, diffusing capacity for carbon monoxide; FEV₁, forced expiratory volume in 1 second; PVD, peripheral vascular disease. *There are no missing data reported because data were collected in a check box (check all that apply) format.

(91%) and mostly men (82%). The majority had significant medical comorbidities. Induction therapy was given in almost half of the patients. Major morbidity (defined as reoperation for bleeding [n = 12], anastomotic leak [n = 261], pneumonia [n = 188], reintubation [n = 227], ventilation beyond 48 hours [n = 571], or death [n = 63]) occurred in 553/2315 patients (24%). Overall, 57% of patients (1327/2315) suffered at least 1 adverse event after esophagectomy. Only

TABLE 2. Frequency of intraoperative and postoperative events in patients after esophagectomy for cancer with and without morbidity

Type of event	All patients	Patients without major morbidity	Patients with major morbidity	P value
Pulmonary embolus	0.9% (22/2315)	0.5% (8/1762)	2.5% (14/553)	<.001
DVT	2.0% (47/2315)	1.3% (23/1762)	4.3% (24/553)	<.001
Tracheostomy	4.0% (92/2315)	0.3% (6/1762)	15.6% (86/553)	<.001
Atrial fibrillation	17.5% (406/2315)	14.9% (262/1762)	26.0% (144/553)	<.001
MI	0.7% (15/2315)	0.3% (6/1762)	1.6% (9/553)	<.001
Intraoperative blood transfusion*	7.2% (138/1926)	6.3% (91/1446)	9.8% (47/480)	.131
Postoperative blood transfusion*	15.9% (306/1926)	11.8% (171/1446)	28.1% (135/480)	<.001
RLN paralysis	2.3% (52/2315)	1.5% (27/1762)	4.5% (25/553)	<.001
Renal failure	2.6% (60/2315)	0.7% (13/1762)	8.5% (47/553)	<.001
Sepsis	3/2% (75/2315)	0.7% (12/1762)	11.4% (63/553)	<.001
Chylothorax	1.9% (45/2315)	1.9% (33/1762)	2.2% (12/553)	.988

DVT, Deep venous thrombosis; MI, myocardial infarction; RLN, recurrent laryngeal nerve. *Only available in STS data version 2.07.

a small number of patients required a return to the operating room to manage an adverse event (168/2315, 7.3%) but the majority of those with major morbidity did (97/553, 57.7%; $P < .0001$ when compared with those without major morbidity). The hospital discharge mortality was 2.7% (63/2315). The hospital discharge mortality in those with major morbidity was 11% (63/553), whereas it was 0 in those without major morbidity ($P < .0001$). The overall median length of stay was 10 days, and the mean was 14 days. The mean length of stay was 10.6 days in patients without major morbidity and 25.6 days in those with major morbidity ($P < .0001$).

The univariate associations between patient characteristics and the end point of mortality or major morbidity are seen in Table 1. Age, medical comorbidities, and steroid use are predictors of major morbidity and mortality. In contrast, induction therapy was not statistically associated with major morbidity. Both cigarette use and significant chronic obstructive pulmonary disease (COPD; FEV₁ < 60%) were associated with major morbidity. Race and the Zubrod score were associated with a trend in major morbidity and mortality.

Intraoperative blood transfusion only occurred in 6% (138/2315) of patients and was associated with a trend in morbidity in univariate analysis ($P = .11$). Operative approach was recorded as transhiatal in 415, thoracoabdominal in 139, Ivor-Lewis in 215, thoracotomy-only in 85, 3-hole (McKeown) in 45, and minimally invasive in 89. Operative approach could not be determined in 1327 patients (previous versions of the database did not collect operative approach).

Table 2 details other intraoperative and postoperative events in patients after esophagectomy with and without major morbidity. Important postoperative events, beyond what were included in the morbidity model, were more common in those with major morbidity. These included the need for postoperative blood transfusion, recurrent laryngeal nerve paralysis, renal failure, and sepsis.

The multivariable association between preoperative patient factors and the end point of mortality or major morbidity is

summarized in Table 3. The bootstrap-adjusted C-statistic was 0.621. Statistically significant predictors included age (75 vs 55), congestive heart failure (CHF), coronary artery disease (CAD), peripheral vascular disease (PVD), hypertension, American Society of Anesthesiology (ASA) rating, insulin-requiring diabetes, steroid use, and cigarette smoking.

TABLE 3. Predictors of major morbidity after esophagectomy for cancer

Variable	Odds ratio		P value
	Estimate	95% CI	
Age (y)			
65 vs 55	1.04	0.90–1.20	.593
75 vs 55	1.24	1.07–1.45	.005
Female	1.20	0.92–1.55	.177
Black race	1.76	0.93–3.34	.082
CHF	2.3	1.18–4.49	.015
CAD	1.31	1.05–1.65	.017
PVD	1.55	1.12–2.14	.009
Zubrod score			
1 vs 0	1.13	0.98–1.30	.100
2 vs 0	1.27	0.95–1.69	.100
3 vs 0	1.43	0.93–2.20	.100
4 vs 0	1.62	0.91–2.86	.100
ASA class			
2 vs 1	1.26	1.10–1.46	.001
3 vs 1	1.60	1.20–2.13	.001
4 vs 1	2.02	1.32–3.10	.001
5 vs 1	2.56	1.45–4.52	.001
Insulin diabetes	1.19	1.05–1.36	.009
Hypertension	1.16	1.01–1.32	.029
Steroids	1.81	1.07–3.06	.026
Renal dysfunction	0.95	0.55–1.64	.846
Induction therapy	0.93	0.77–1.11	.424
Cigarette usage	1.27	1.03–1.56	.022
BMI (per 5-unit increase)	1.02	1.00–1.03	.123
Time trend (per 5 y)	1.29	0.93–1.80	.133

ASA, American Society of Anesthesiology; BMI, body mass index; CAD, coronary artery disease; CHF, congestive heart failure; CI, confidence interval; PVD, peripheral vascular disease.

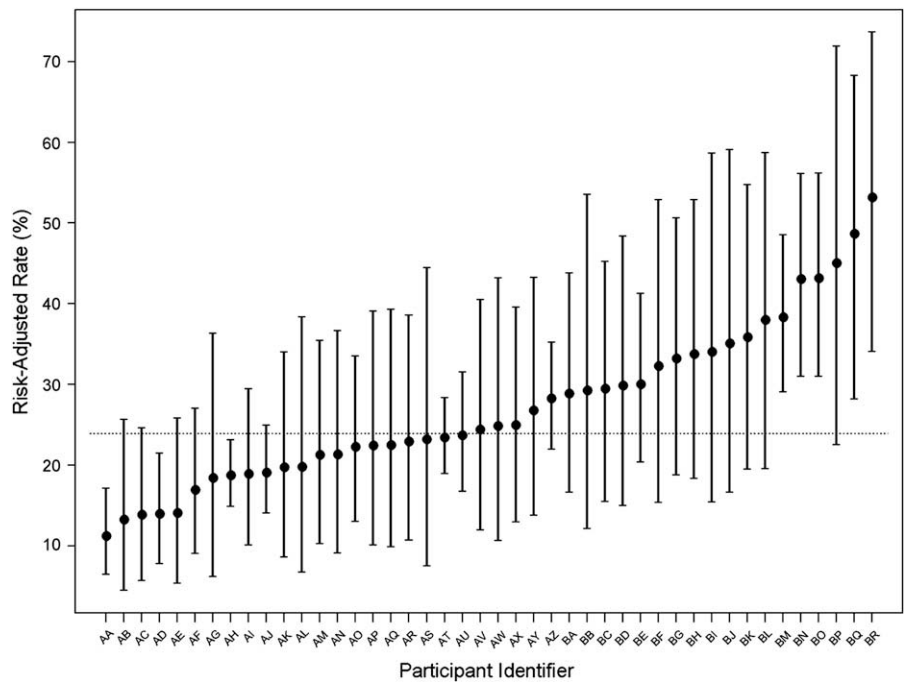


FIGURE 1. Risk-adjusted rates of morbidity after esophagectomy for cancer among Society of Thoracic Surgeons sites. The graph excludes participants that contributed fewer than 6 operations to the analysis data set. The participant identifiers are anonymous codes and do not denote an abbreviation for a center’s name.

The regression coefficients for this model are presented in the [Appendix](#). When average annual volume was included in the regression analysis as a linear covariate, the odds ratio estimate for a 10-unit decrease in volume was 1.09 (95% confidence interval 0.98–1.20; $P = .10$).

[Figure 1](#) examines the hospital-specific risk-adjusted rates of mortality and morbidity among the 44 sites that had at

least 1 esophagectomy per year (29 sites had <6 esophagectomies in the 6-year study period and were excluded from this particular analysis). The probability intervals of some of the best-performing sites (on the left side) do not overlap with some of the sites with worse outcomes (on the right side), indicating there is a significant difference in performance among some sites.

The relationship between major morbidity and volume among the same 44 sites is plotted in [Figure 2](#). The average annual case volume at these 44 sites ranged from 1 to 83 in the 6-year reporting period. Note there is no strong volume performance relationship in the graph, a finding that is consistent with the nonsignificant volume effect in our multivariable regression analysis.

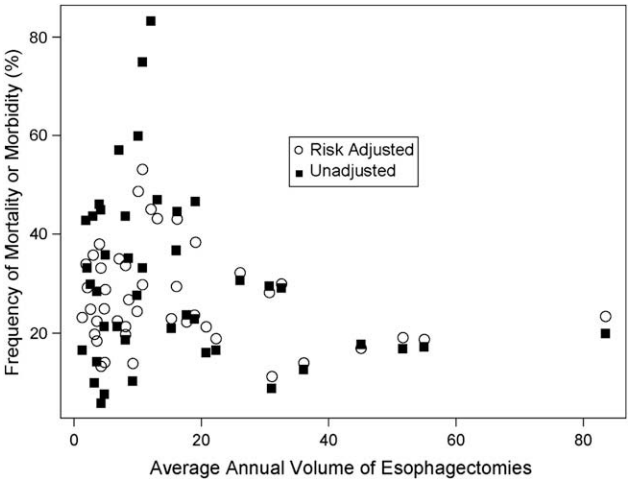


FIGURE 2. Comparison of adjusted and unadjusted morbidity rates versus participant average annual volume of esophagectomies for lung cancer. Each dot and square represent 1 participant site. The graph excludes participants that contributed fewer than 6 operations to the analysis data set.

DISCUSSION

STS sites that participate in the GTDB perform esophagectomy for cancer with relatively low mortality and morbidity. Patients who experience adverse events after esophagectomy have a marked increase in length of stay and mortality. We identified multiple risk factors for combined morbidity and mortality after esophagectomy for cancer: age, CHF, CAD, PVD, diabetes, hypertension, use of steroids, smoking status, and ASA score. Using these results, we constructed a risk model for esophagectomy that allows individual STS sites to compare their results with others as a means toward quality improvement. In contrast to previous studies using administrative data,^{9,10} we did

not identify a strong association between volume and the composite measure of morbidity and mortality among STS sites that participate in the GTDB.

The operative mortality among sites that participate in the STS GTDB was 2.7%, which is markedly lower than recent reports based on large nationwide data sets.^{8,12,13} A recent report from the Department of Veterans Affairs National Surgical Quality Improvement Program (VA NSQIP) recorded a mortality of 9.8% among 1777 patients who had esophagectomy.⁸ Another report based on the national Surveillance, Epidemiology, and End Results (SEER) Medicare-linked database recorded a mortality of 11% among 868 patients who had esophagectomy.¹² Additionally, the average length of stay (14 days) in our patient cohort was markedly shorter than the average length of stay (21 days) in the SEER patient cohort. These differences may reflect overall better quality provided by surgeons who are board certified and participate in a voluntary quality improvement initiative.

Several factors likely contribute to the lower observed morbidity and mortality of the STS GTDB. Participation is encouraged but not mandated in the STS GTDB, which leads to selection bias regarding participants. As such, early reporters to the GTDB are likely to be clinicians with a particular interest in quality improvement. Second, given the early academic nature of the database, participants possibly overrepresent academic medical centers. In addition, academic medical centers typically have more resources to support ill patients and experience with complex procedures such as esophagectomies. Last, the database was initially limited to board-certified thoracic surgeons, who have been shown to have better outcomes after esophagectomy.¹⁴ The results achieved within the context of the STS GTDB are clearly superior to those reported for the entire US population and are hence not representative of the entire esophageal cancer patient population. It is likely that the “best” surgeons are overrepresented within this data set.

Our analysis identified 8 predictors of morbidity/mortality following esophagectomy for cancer. Age is an intuitive risk factor and has been previously identified as an important predictor of outcome following esophagectomy.^{1,2} A large single-institution report from a center of excellence, however, has suggested that age per se should not be a limiting criteria for esophagectomy.¹⁵ This report suggests that the use of stringent patient selection criteria in elderly patients can mitigate their increased risk, a conclusion with which we agree. CHF was found to be a third important predictor but was rarely present among our patients (51/2315, 2.2%), likely due to careful case selection. Given the large fluid shifts that occur after esophagectomy, it is not surprising that CHF is a risk factor. Careful patient selection, preoperative medical optimization, and conscientious postoperative care could all improve outcomes in these patients. PVD has not been well studied, but a recent report using the

Nationwide Inpatient Sample (NIS) database found PVD to be associated with increased risk.¹³ In our data set, PVD was again quite uncommon, occurring in only 5.7% (132/2315) of our patients, but was strongly associated with increased morbidity. It is likely that the end-stage effects of both smoking and unfavorable cardiovascular biology lead to worse outcomes. Proper patient selection and medical optimization of cardiac risk would seem to be important strategies to minimize perioperative risk. Hypertension and CAD are other intuitive risk factors in this patient population and have not been reported before.

Insulin-requiring diabetes has been previously identified by the VA NSQIP model and other reports as a risk factor for both morbidity and mortality following esophagectomy for cancer.^{8,16} Diabetic patients requiring insulin were again uncommon in our study population, representing only 3.7% (86/2315) of our patients. Again, proper preoperative medical optimization and careful diabetic management after operation likely lead to optimal outcomes. A history of smoking was present in the majority of our patients (1731/2315, 75%) and was associated with worse outcomes. We were not able to tease out outcome differences depending on when or if patients stopped smoking. Certainly active smoking should be strongly discouraged and patients referred to smoking cessation counseling well before an esophagectomy. Others have reported that significant COPD (either the presence of the diagnosis or a low FEV₁) is associated with increased morbidity and mortality after esophagectomy.^{2,8,17} Most of our patients did not have FEV₁ recorded (we are unsure if this was an oversight or if it was not performed as part of preoperative testing) but among those who did (923/2315, 40%), it was highly predictive of postoperative morbidity in univariate analysis. Given that pulmonary complications are the most common major complications after esophagectomy, and in particular the complications most likely to lead to death, it is intuitive that significant COPD leads to worse outcomes. We need to encourage surgeons to increase the frequency of ordering pulmonary spirometry prior to esophagectomy to both better risk-stratify patients and to identify patients who might benefit from preoperative pulmonary optimization.

The ASA rating, a surrogate for medical comorbidities, was also strongly associated with morbidity. More complex patients (based on other complexity scoring systems) have been previously reported as more likely to have increased morbidity after esophagectomy.^{1,8}

An important risk factor for combined morbidity and mortality, black race, has not been evident in many other large multi-institutional database reports.^{8,12} One recent report from the NIS database, however, concurs with our data; in this data set, blacks had an increased risk of mortality after esophagectomy.¹³ As African-Americans constituted only 2.7% (63/2315) of our patient population, the validity of our observation needs to be confirmed by other larger data

sets. It is possible that other unmeasured comorbidities, observed with greater frequency in the African-American population, will ultimately explain this increased risk. The issue of obesity is particularly relevant now that the majority of patients presenting with esophageal cancer are overweight and have Barrett's adenocarcinoma from prolonged reflux. The mean body mass index (BMI) in our patients was 28 (overweight, close to obese). We found obesity to be associated with a trend in increased morbidity. Previous reports are conflicting about obesity, with some suggesting it is an important risk factor and others not.^{13,17,18} The largest report suggests that obesity is associated with more specific complications (such as blood loss and recurrent laryngeal nerve injury) but is not associated with increased mortality.¹⁸ This report is partially limited by the fact that all operations were transhiatal esophagectomies. It is intuitive that the very obese would have longer, more difficult operations and have impaired mobility and pulmonary function early after operation. Further investigation into the type of operative approach (transhiatal versus a thoracotomy) is warranted to ascertain if operative risk in the very obese might be mitigated by a nonthoracotomy approach.

We did not find induction therapy to be predictive of increased major morbidity. Some previous reports suggest an increased risk, and others do not.⁵⁻⁸ To date, our report is by far the largest and most geographically representative of the entire United States; however, it still suffers from the previously mentioned potential bias that STS GTDB participants may be select high-performing clinicians not fully representative of all surgeons in the United States. Regardless, our data suggest that careful patient selection and management can allow this therapy to be given if necessary without increased morbidity.

Importantly, we did not find a strong volume performance relationship for the composite end point of major morbidity and mortality after esophagectomy. Figure 2 does not show a strong relationship between volume and composite mortality/morbidity despite a wide variation in average annual volume (1 to 83 patients per year). Numerous previous reports based on administrative data sets have demonstrated a volume performance relationship in esophagectomy for mortality.^{9,10} As previously noted, several reports also indicate a volume performance relationship with postoperative morbidity as well, which is not surprising given that complications are what lead to mortality.^{19,20} Many have even suggested using volume as a quality surrogate for esophagectomy (eg, the Leapfrog Group) to direct referrals to appropriate surgeons. Our analysis, using a composite end point of mortality or major morbidity, does not support this approach, at least as applied to surgeons who participate in the STS GTDB. This likely reflects the skill of surgeons at such centers and the overall level of supportive care available to assist in the perioperative management of these complex patients. It must be acknowledged that

these results cannot necessarily be extrapolated to other scenarios, such as non-Board-certified thoracic surgeons who do not participate in the STS GTDB. It is important to emphasize that we did not study mortality alone like previous reports on the volume performance relationship. We can state, however, that the volume performance relationship seems to be partially mitigated among participants in the STS GTDB with a composite measure of morbidity and mortality.

There are several limitations of our report. Most importantly, this database represents a select group of surgeons from within the STS who are interested in quality improvement in an altruistic fashion. There is nothing financial to be gained from database participation, and in fact it is relatively expensive to participate in it. The surgeons who participate currently are likely to be "first adopters" and hence are probably not representative of the entire US surgical community. The database is currently not audited for data quality, although there are plans in the future for this to occur. However, there is no reward for any potential "gaming" the system, so it is unlikely that this is a systematic problem. There were more missing data than we would have liked, especially in the staging and pulmonary function data. We did analyze our data set with and without these data and there was no substantial difference in results, however. We did not collect data with regard to rehospitalization related to the esophagectomy, so it is likely we missed some late complications of esophagectomy. Last, we analyzed in-hospital mortality, so it is possible that we missed some early deaths that occurred after hospital discharge.

We conclude that thoracic surgeons participating in the STS GTDB perform esophagectomy with a low mortality. Age, medical comorbidities, smoking status, and significant obstructive lung disease are predictors of major morbidity and mortality after esophagectomy for esophageal cancer. Prognostic factors identified in this analysis will help to predict risk in individual patients and guide quality improvement through risk-adjusted feedback. Volume alone is an inadequate proxy for quality assessment after esophagectomy. Modern multi-institutional prospectively collected clinical data sets with sophisticated risk-adjustment methodology are required to compare results.

References

1. Atkins BZ, Shah AS, Hutcheson KA, et al. Reducing hospital morbidity and mortality following esophagectomy. *Ann Thorac Surg*. 2004;78:1170-6.
2. Ferguson MK, Durkin AE. Preoperative prediction of the risk of pulmonary complications after esophagectomy for cancer. *J Thorac Cardiovasc Surg*. 2002;123:661-9.
3. Ferguson MK, Martin TR, Reeder LB, Olak J. Mortality after esophagectomy: risk factor analysis. *World J Surg*. 1997;21:599-604.
4. Visbal AL, Allen MS, Miller DL, et al. Ivor Lewis esophagogastrrectomy for esophageal cancer. *Ann Thorac Surg*. 2001;71:1803-8.
5. Rice DC, Correa AM, Vaporciyan AA, et al. Preoperative chemoradiotherapy prior to esophagectomy in elderly patients is not associated with increased morbidity. *Ann Thorac Surg*. 2005;79:391-7.

6. Reynolds JV, Ravi N, Hollywood D, et al. Neoadjuvant chemoradiation may increase the risk of respiratory complications and sepsis after transthoracic esophagectomy. *J Thorac Cardiovasc Surg.* 2006;132:549-55.
7. Lin FC, Durkin AE, Ferguson MK. Induction therapy does not increase surgical morbidity after esophagectomy for cancer. *Ann Thorac Surg.* 2004;78:1783-9.
8. Bailey SH, Bull DA, Harpole DH, et al. Outcomes after esophagectomy: a ten year prospective cohort. *Ann Thorac Surg.* 2003;75:217-22.
9. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med.* 2002;346:1128-37.
10. Chang AC, Birkmeyer JD. The volume performance relationship in esophagectomy. *Thorac Surg Clin.* 2006;18:87-94.
11. Society of Thoracic Surgeons 2008. Available at: www.sts.org/sections/stsnationaldatabase. Accessed March 24, 2008.
12. Chang AC, Ji H, Birkmeyer NJ, et al. Outcomes after transhiatal and transthoracic esophagectomy for cancer. *Ann Thorac Surg.* 2008;85:424-9.
13. Rodgers M, Jobe BA, O'Rourke RW, et al. Case volume as a predictor of inpatient mortality after esophagectomy. *Arch Surg.* 2007;142:829-39.
14. Dimick JB, Goodney PP, Orringer MB, et al. Specialty training and mortality after esophageal cancer resection. *Ann Thorac Surg.* 2005;80:282-6.
15. Ruol A, Portale G, Zaninotto G, et al. Results of esophagectomy for esophageal cancer in elderly patients: age has little influence on outcome and survival. *J Thorac Cardiovasc Surg.* 2007;133:1186-92.
16. Karl RC, Schreiber R, Boulware D, et al. Factors affecting morbidity, mortality and survival in patients undergoing Ivor Lewis esophagogastrectomy. *Ann Surg.* 2000;231:635-43.
17. Avendano CE, Flume PA, Silvestri GA, et al. Pulmonary complications after esophagectomy. *Ann Thorac Surg.* 2002;73:922-6.
18. Scipione CN, Chang AC, Pickens A, et al. Transhiatal esophagectomy in the profoundly obese: implications and experience. *Ann Thorac Surg.* 2007;84:376-83.
19. Dimick JB, Pronovost PJ, Cowan JA, et al. Variation in postoperative complication rates after high-risk surgery in the United States. *Surgery.* 2003;134:534-41.
20. Wouters MW, Wijnhoven BP, Karim-Kos HE, et al. High-volume versus low-volume for esophageal resections for cancer: the essential role of case mix adjustments based on the clinical data. *Ann Surg Oncol.* 2008;15:80-7.

Appendix

Covariates

The following patient-level covariates were included in the multivariable analysis: age (modeled as a 2-phase linear function with a change of slope at 65 years); gender (female/male); race (African-American/non-African-American); CHF (yes/no); CAD (yes/no); PVD (yes/no); Zubrod Score (5 categories, 0-4, modeled as linear); ASA class (5 categories, 1-5, modeled as linear); insulin-dependent diabetes (yes/no), hypertension (yes/no), steroid use (yes/no), renal dysfunction (creatinine > 2 or dialysis/none), induction therapy (chemotherapy or thoracic radiation/none), smoking (ever/never), BMI (linear), year of surgery (linear). These covariates were selected based on a combination of a priori clinical judgment and an empirical investigation of missing data frequency. Variables were not added or removed from models based on an automated variable selection procedure. In addition to patient factors, all models were estimated with and without the inclusion of a linear term for the effect of hospital volume of esophagectomy procedures.

Ordinary Logistic Regression Model

The relationship between patient predictor variables and morbidity was examined first using ordinary logistic regression and subsequently using hierarchical logistic regression. The ordinary logistic model has the form:

$$\log(p_i/[1-p_i]) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_q x_{qi}$$

where p_i denotes the probability of major morbidity for the i th patient; x_{qi} denotes the value of q th covariate for the i th patient, and β_1, \dots, β_q denote unknown parameters to be estimated from the data. The term x_{qi} represents quantitative risk factors such as age and binary indicator variables (eg, 1 = male, 0 = female). Parameters of the ordinary logistic regression model were estimated using generalized estimating equations methodology. This approach results in estimates that are similar to conventional maximum likelihood estimation, but the standard errors are adjusted to account

for statistical dependence (clustering) between outcomes of patients at the same hospital.

Hierarchical Logistic Regression Model

To study between-hospital variation in risk-adjusted outcomes, the model described above was subsequently refit as a 2-level hierarchical model with patients nested within participants. Unlike ordinary logistic regression, this model included a set of hospital-specific random effects. The form of the hierarchical logistic model was:

$$\log(p_{ji}/[1-p_{ji}]) = \beta_0 + \beta_1 x_{1ji} + \dots + \beta_q x_{qji} + e_j$$

where p_{ji} denotes the probability of major morbidity for the i th patient at the j th hospital; the x_{qji} denotes the covariate values for the i th patient at the j th hospital; and e_j denotes a (random effect) intercept parameter for participant j . Parameters of the random effects logistic model were estimated in a Bayesian framework using WinBUGS software. In a Bayesian analysis, inferences about unknown quantities are expressed in terms of probabilities. For example, it is possible to report the probability that the coefficient for a given variable is larger than 0 or the probability that a hospital's random effect parameter is greater than 0.

Hospital-specific RARs of mortality or major morbidity were calculated using the estimated hospital-specific random effects parameters. The risk-adjusted rate is a prediction of what a given hospital's actual morbidity rate would be if all patients in the STS database received surgery at that hospital. The risk-adjusted rate of the h th hospital, denoted RAR _{h} , is calculated by the following formula:

$$\text{RAR}_h = \frac{1}{N} \sum_{j=1}^{73} \sum_{i=1}^{n_j} \frac{e^{\beta_0 + \beta_1 x_{1ji} + \dots + \beta_q x_{qji} + e_h}}{1 + e^{\beta_0 + \beta_1 x_{1ji} + \dots + \beta_q x_{qji} + e_h}}$$

where n_j denotes the number of patients at hospital j , and $N = n_1 + n_2 + \dots + n_q$. The subscript h in e_h denotes that the random effect for the h th hospital is held constant in the summation

and is substituted into the probability formula for all patients in all hospitals. Estimated RAR values were plotted along with a 95% Bayesian probability intervals. The Bayesian probability interval has the interpretation that it is 95% likely that the true risk-adjusted rate lies between the upper and lower limits of the reported interval. If the probability interval includes the STS average rate, then the hospital's performance is not statistically distinguishable from average.

Missing Data

Missing data methods for ordinary logistic model. Missing predictor values in the ordinary logistic model were managed using a combination of complete case analysis and single imputation. For age, gender, and year of surgery, patients with missing data were excluded from the analysis (ie, complete case analysis). This approach was adopted because nonmissing data on these 3 fields is a basic inclusion criterion for all reports produced by the STS data warehouse. For comorbidity variables (CHF, CAD, insulin diabetes, hypertension, steroid use, renal dysfunction, induction therapy, smoking), missing values were defaulted to the negative (ie, single imputation). The decision to use single imputation was based on consideration of the STS data collection instrument. Comorbidities in STS are collected using a check box format. Because no data entry is required unless a comorbidity is present, there is no opportunity to distinguish the absence of a comorbidity from missing data. Thus, when data are received into the warehouse that contain a null value, it is not clear whether this null value indicates missing data or the absence of a comorbidity, and in fact this may depend on the software that was used for collecting the data. Because missing data are not distinguishable from null values, the only option for analysis is to model yes versus no or missing, which is equivalent to imputing missing data to the negative. For the remaining variables, missing data were imputed to the median (BMI) or mode (race, Zubrod score, ASA class.)

Missing data methods for hierarchical model. To explore whether results change depending on the method of handling missing data, 2 approaches of handling missing data were implemented and compared to one another. In both approaches, age, gender, year of surgery, and comorbidities were imputed using the approach described above. For the remaining variables (race, Zubrod score, ASA class, BMI), missing data were either imputed using the method described above (approach 1) or were imputed using a Bayesian multiple imputation model (approach 2). To implement approach 2, the probability distribution of race, Zubrod score, ASA class, and BMI was approximated as a multivariate normal distribution, and the mean of each variable was modeled as a linear function of the following variables: age, CHF, hypertension, induction therapy, cigarette usage, year of surgery, and presence of postoperative major morbidity. Parameters of the imputation model were

estimated jointly (in the same Markov Chain Monte Carlo procedure) with the parameters of the model predicting major morbidity. At each iteration, covariate values for patients with missing data were randomly sampled from the posterior predictive distribution and substituted into the model for the relationship between covariates and morbidity. Because the imputation model uses a multivariate normal distribution, the imputed values are continuous rather than categorical. Following the advice of Allison and others, the imputed versions of race, Zubrod score, and ASA class were left as continuous variables and were not rounded to integers. The final model results were obtained by averaging across simulation iterations. The impact of the 2 missing data methods was assessed by comparing odds ratios obtained under approach 1 versus approach 2. For each covariate, the estimated odds ratio changed by less than 3% (on a relative scale) depending on the choice between approach 1 and approach 2. Because the results were fairly insensitive to the choice of missing data method, the simpler method (approach 1) was adopted.

Model Coefficients

Variables and regression coefficients for the model in Table 3 are summarized in Table A1.

TABLE A1. Variables and regression coefficients for the logistic regression model reported in Table 3

Model variable	Coefficient	Definition of model variable
Constant	-2.8341	Constant
Age function 1	0.0179	= max(65-age, 0)
Age function 2	-0.0039	= max(age-65, 0)
Female	0.1789	= 1 if patient is female, = 0 otherwise
Black	0.5660	= 1 if patient is black, = 0 otherwise
CHF	0.8332	= 1 if patient has CHF, = 0 otherwise
CAD	0.2737	= 1 if patient has CAD, = 0 otherwise
PVD	0.4359	= 1 if patient has PVD, = 0 otherwise
Zubrod	0.1199	Ordinal variable taking the values 0, 1, 2, 3, 4
ASA	0.2350	Ordinal variable taking the values 1, 2, 3, 4, 5
Insulin diabetes	0.1754	= 1 if patient has insulin-dependent diabetes, = 0 otherwise
Hypertension	0.1472	= 1 if patient has hypertension, = 0 otherwise
Steroids	0.5944	= 1 if history of steroid usage, = 0 otherwise
Renal dysfunction	-0.0544	= 1 if patient has creatinine > 2.0 or dialysis, = 0 otherwise
Induction Tx	-0.0744	= 1 if chemotherapy or radiation preoperatively, = 0 otherwise
Cigarettes	0.2381	= 1 if history of smoking, = 0 otherwise
BMI	0.0150	Continuous variable
Year*	0.0511	= Year of surgery minus 2002

ASA, American Society of Anesthesiology; BMI, body mass index; CAD, coronary artery disease; CHF, congestive heart failure; PVD, peripheral vascular disease; Tx, treatment. *To avoid extrapolation, future model estimates should be calculated as if the year of surgery was 2007.

Discussion

Dr J. Luketich (Pittsburgh, Pa). Dr Wright, that was an excellent presentation, and I congratulate you and your colleagues on your efforts to utilize the STS database, which includes close to 2000 cases of esophagectomy, to create a model of perioperative risk. One very important finding of your study, as you point out, is that thoracic surgeons participating in the STS database have a markedly lower mortality rate of only 2.5% compared with the Medicare database showing alarming mortality rates between 8% and 23%. I have several questions about your study.

What is the quality control of the data collection of the STS database? Are charts periodically audited by the STS during site visits? Along these lines, has your statistician voiced concerns about including the pulmonary function data, with fewer than 40% of patients having reported spirometry yet the finding of an overall correlation with outcome? And I wonder if you could comment specifically on which morbidities that the lower FEV₁ was associated with. And has this influenced your practice? If you encounter a patient with an FEV₁ < 60%, are you excluding them from esophagectomy?

Dr Wright (Boston, Mass). Thank you for that question. Currently, the STS does not audit the data in their database. This is a new database, in existence for only 6 years. You have to remember that the cardiac database, in existence for over 15 years, has only been audited since 3 years ago, I believe. Their initial audit, I think, was recognized as a success, indicating a less than 10% variation in data elements. We anticipate the same results, but we have not started that yet. It is clearly something we need to do.

We specifically excluded the use of pulmonary function tests in our multivariate model because only 40% of our patients had pulmonary function tests performed. Thus we only did a univariate analysis of pulmonary function tests. I think that was an honest way to treat that as we had so much missing data. Certainly, we encourage surgeons to enter these data, because this is clearly going to be an important risk factor.

Dr J. Luketich. Thank you. Second question. How serious was the missing data problem with other variables? And I think you have answered the other questions about the auditing plans. But was the rate of missing variables of key comorbidity and outcome variables of concern?

Dr Wright. Well, missing variables are always of concern to the people looking at the database, but I think they were within reason. For race, age, and gender, there were no missing variables. BMI was 13%; cigarette smoking, 13%; diabetes, 5%; peripheral vascular disease, 13%. The rate of missing variables for outcome measures was <3%.

Dr. Luketich. My final question is in regards to the lack of a volume/outcome relationship for the number of esophagectomies performed annually in each center. Your analysis included 40 of the 68 sites, and as I understand it, 28 sites performing fewer than 5 esophagectomies per year were excluded from the analysis. I wonder if the results would be similar if all sites were included? It would seem like the hospitals with the very lowest numbers of esophagectomies annually are the very hospitals we want to examine when it comes to low volume of index cases and high complication and death rates. Did you analyze and correct for other factors such as surgeon volume and specialty training? And do you think that the requirement for thoracic board certification and performance of esophagectomies in academic medical centers influence your findings?

Dr Wright. That is an excellent question, a 2-part question. In answer to the first part, our statisticians thought it was not statistically valid to include sites that had fewer than 1 esophagectomy per year, because they were looking on a per-year basis. It would be interesting to go back and see what those results were. But certainly the sites to the right on my graph show 1 case a year versus 60 cases a year. There is not a lot of difference.

Certainly I don't propose that this volume/performance relationship is true of all of America. The STS thoracic database participants are very select. They belong to the STS, they are board certified, and they are very interested in quality improvement, and I think that is why the results are so good.

Dr S. DeMeester (Los Angeles, Calif). Cam, thank you very much for that interesting presentation and data. Just a quick question. Did you analyze induction therapy as either yes or no, they had it or didn't have it, or did you stratify by the amount of radiation? As you know, it has been demonstrated that high radiation doses seem to be associated with the risk of increased morbidity or mortality.

Dr Wright. That is another good question, and we currently in our data field only collect radiation as yes or no. So we could not do that stratification.

Dr J. Benfield (Los Angeles, Calif). Many esophagectomies are done by general surgeons. Do you see any way to gather meaningful data to compare outcomes of general surgeons with those of thoracic surgeons?

Dr Wright. Well, I believe that paper has already been written by Mark Orringer's group at the University of Michigan, and they did show that there was a modest improvement in results if you were a board-certified thoracic surgeon as opposed to a general surgeon.

Dr T. Karamlou (Portland, Ore). Two quick questions. One, what was the conduct of the esophagectomies? Were these transhiatal, transthoracic, Ivor-Lewis, laparoscopic, 3 fields? And number 2, how is the diagnosis of some of the comorbidities arrived at, specifically pneumonia? Was this just a yes/no, or were there strict criteria for establishing this as you used it as one of your composite end points?

Dr Wright. Pneumonia is defined in the database. It is a standard definition of a white count, fever, and a change in sputum. All the definitions are standardized. We did collect what type of esophagectomy was done, and transhiatal was by far and away the commonest esophagectomy performed. But all varieties were looked at, including Ivor-Lewis and the 3-hole type. We did not stratify outcome according to esophagectomy type.

Dr N. Altorki (New York, NY). Cam, I enjoyed your presentation. I was wondering why you did not include in the measure of morbidity electrical instability, atrial arrhythmias, and this sort. In our experience, it has been the cause of major morbidity and prolongation of the hospital stay, and why was that not entered in your model?

Dr. Wright. That was a clinical judgment decision in terms of how major is major. We are aware that atrial fibrillation is a marker for other bad things happening. Most serious events are respiratory events, and so we were really focused on pulmonary complications and morbidity and death after esophagectomy. But I take your point that it is a judgment call.